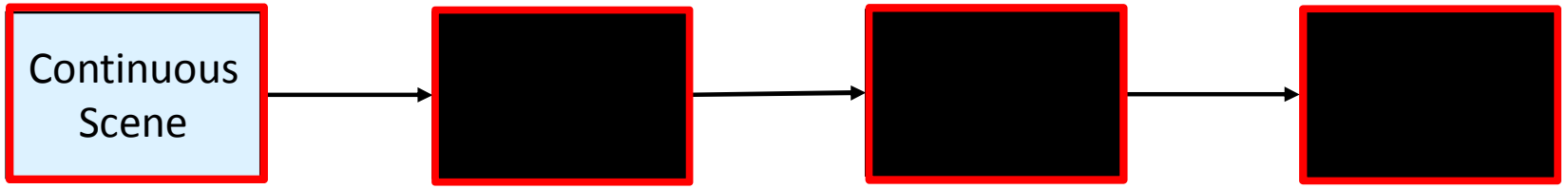
Mahidol University *Wisdom of the Land*

Chapter 14

Video Processing

Video Basics

- Video is any sequence of time varying images.
- In the video, the image information is digitized both temporally and spatially and the resultant pixel gray-levels are quantized.



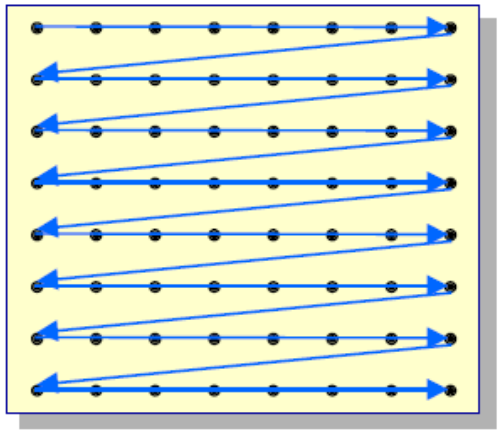
- Temporal sampling
  - The video consists of a sequence of images, displayed in rapid succession, to give an illusion of continuous motion. If the time gap between successive frames is too large, the viewer will observe jerky motion.
  - In practice, most video formats are used temporal sampling rates of 24 frames per second and above.

Video Basics

- Spatial Sampling
  - In the digital representation of the image, the value of each pixel needs to be quantized using some finite precision.
  - In practice, 8 bits are used per luminance sample.

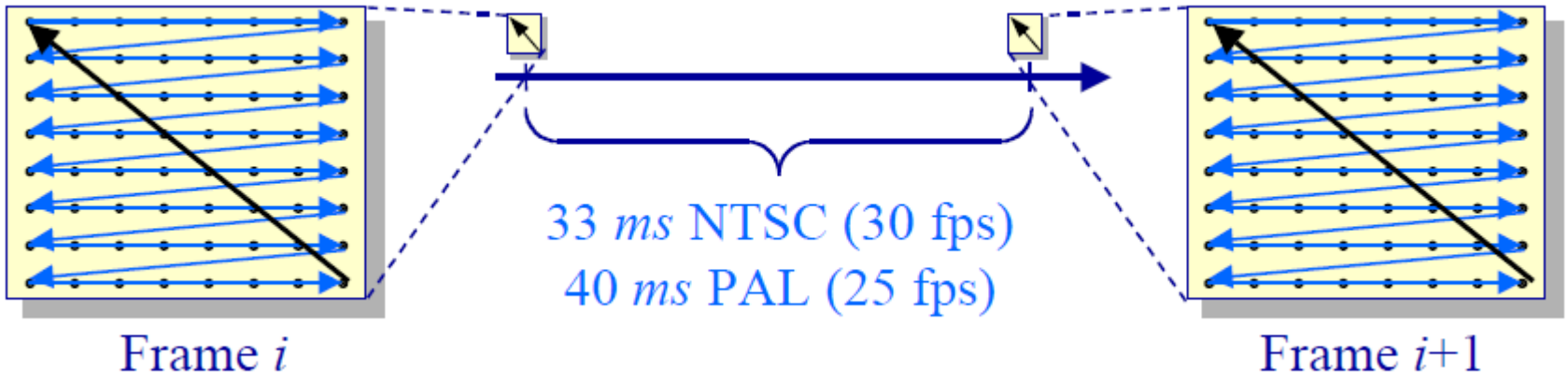
Video Formats

- Representation of a 2-dimensional image



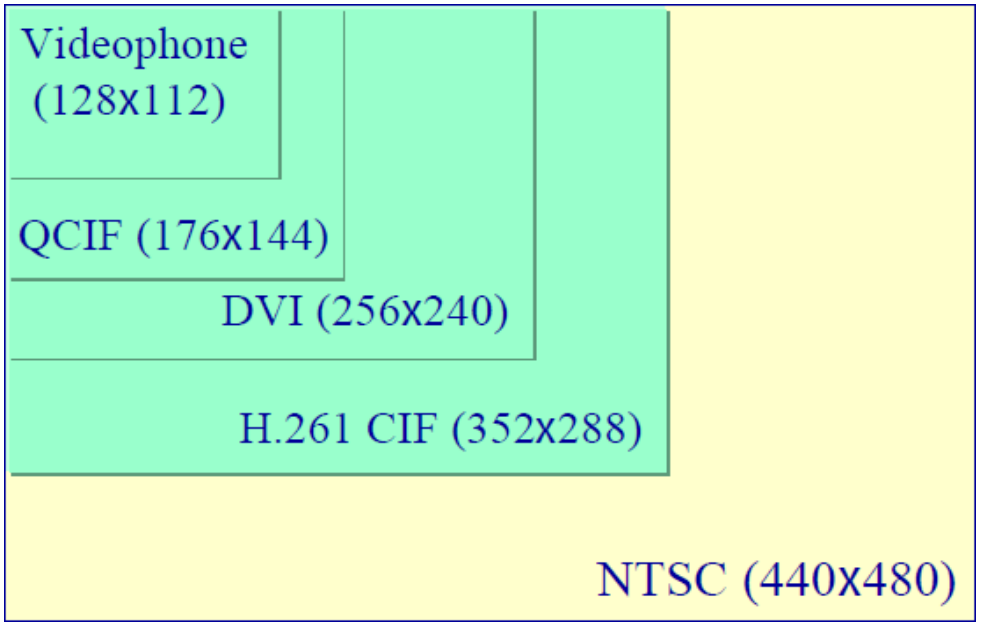
$(R, G, B)_{11}, (R, G, B)_{12}, (R, G, B)_{13}, \ldots, (R, G, B)_{row,\ col}$

- Representation of motion (3-dimensional images)



Video Formats

- Television broadcast standards
  - NTSC — 525 lines
  - PAL — 625 lines
- Computer graphics standards
  - VGA — 640x480
  - SVGA — 1024x768
- Multimedia standards
  - CIF — 352x288
  - QCIF — 176x144
- Digital video standards (dots per lines)
  - SD — 640 × 480 (720p)
  - HD — 1280x720 (720p) / 1920x1080 (1080p)



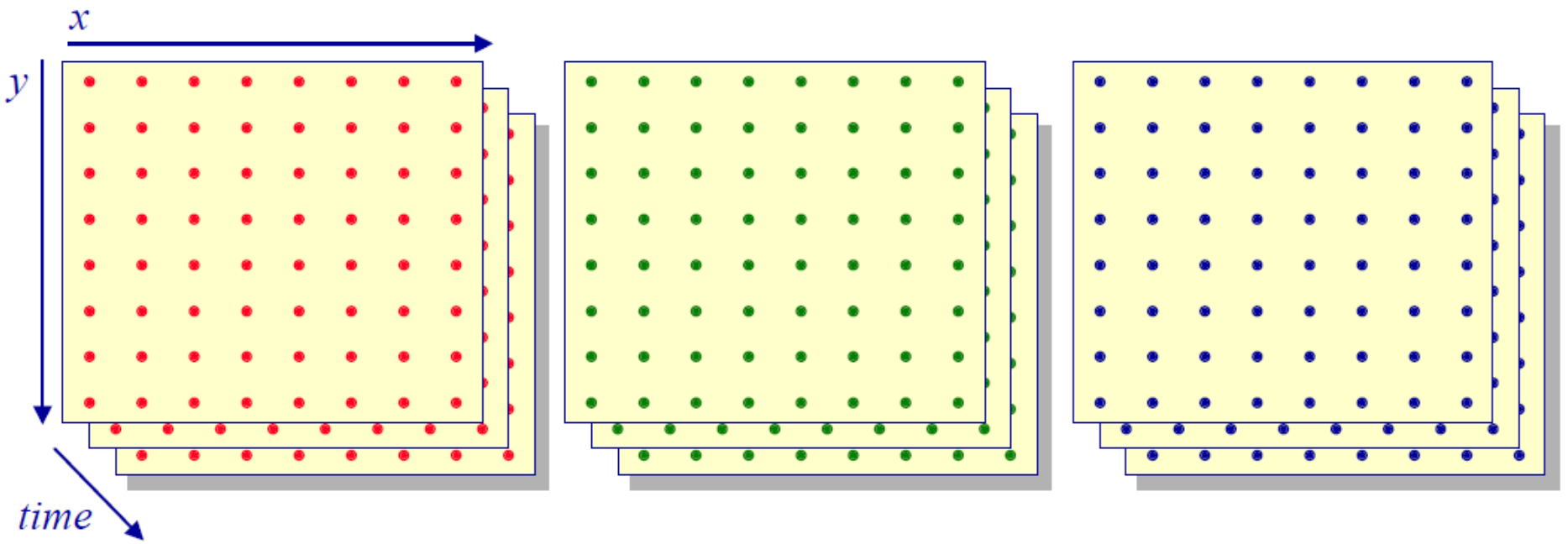
Video Formats

- The common Video Formats
  - Video frames that are displayed at a prescribed frame rate. For example, frame rate of 30 frames/sec is used in National Television System Committee (NTSC) video.
  - For video conferencing, the Common Intermediate Format (CIF) has 352 x 288 pixels, and the Quarter CIF (QCIF) format has 176 x 144 pixels.

Color models

- The components of video transmission
  - Video is a multi-dimensional signal
  - The primary colors in this domain are red, green, and blue.

Red component Green component Blue component

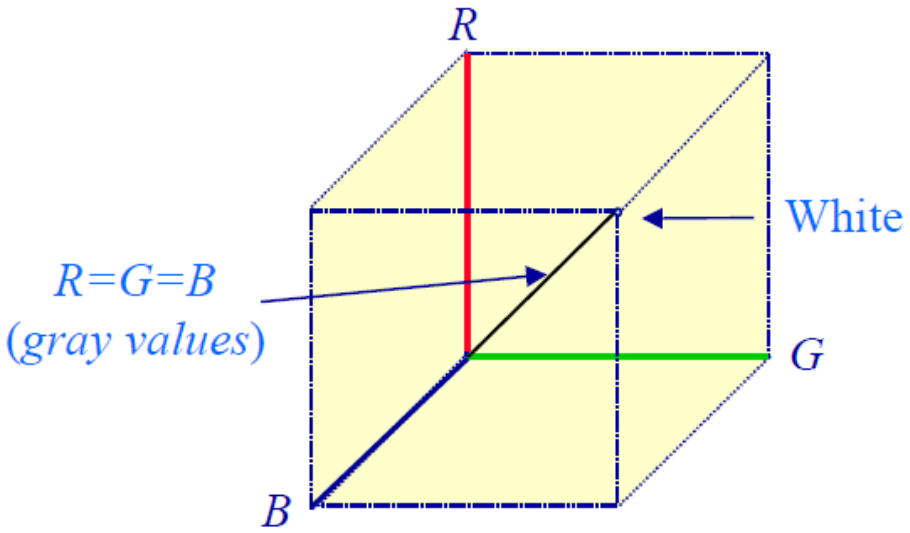


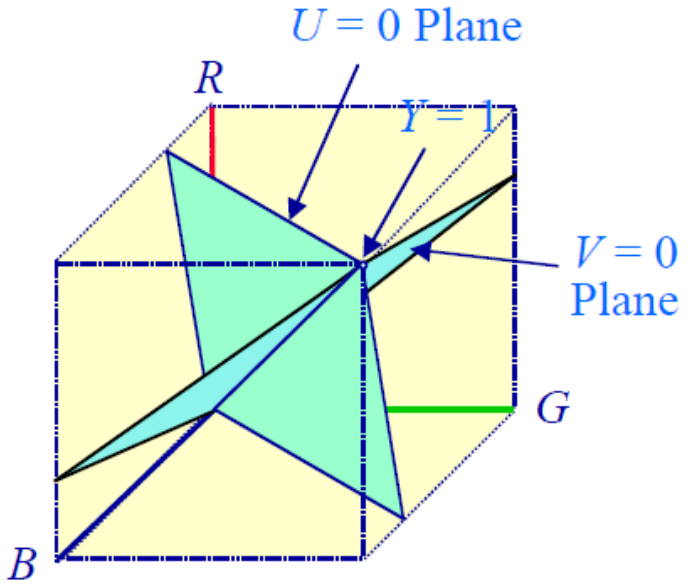
Color models

- RGB is not widely used for transmitting a signal between capture and display devices
  - It is difficult to manage 3 component inputs & outputs that are requires too much bandwidth.

- Composite formats are used instead
  - Luminance (“Y”) — the brightness of the monochrome signal
  - Chrominance — the coloring information
  - Chrominance is typically represented by two “color difference” signals:
    - “U” and “V” (“hue and tint”) or
    - “I” and “Q” (“saturation” and “color”)

Color models





- NTSC video
  - Y = 0.30R + 0.59G + 0.11B
  - I = 0.60R – 0.28G – 0.32B
  - Q = 0.21R – 0.52G + 0.31B

- PAL video/Digital recorders
  - Y = 0.3R + 0.6G + 0.1B
  - U = (B – Y) x 0.493
  - V = (R – Y) x 0.877

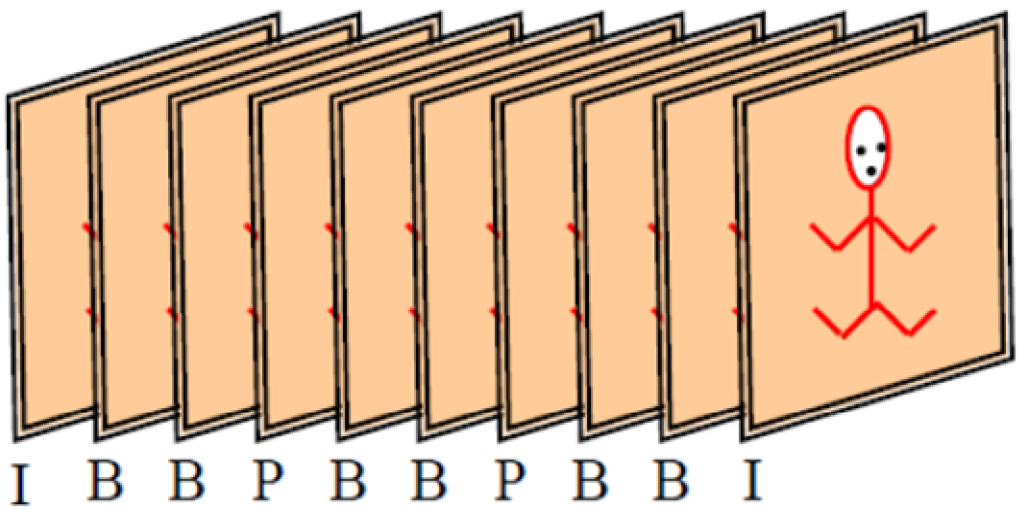
Color models

- Sample an analog representation of video (RGB or YUV) & quantize
  - Two dimensions of video are already discretized
  - Sample in the horizontal direction according to the resolution of the media

- Storage/transmission requirements
  - NTSC — 440 x 480 x 30 x 24 = 152x106 bits/sec (19 MB/s or 24 bits/pixel (bpp))

Video Frame Types

- Three types of video frames are I-frame, P-frame and B-frame.
  - ”I” stands for Intra coded frame
  - “P” stands for Predictive frame
  - “B” stands for Bidirectional predictive frame



Video Frame Types

- “I” frames are encoded without any motion compensation and are used as a reference for future predicted “P” and “B” type frames. “I” frames however require a relatively large number of bits for encoding.
- “P” frames are encoded using motion compensated prediction from a reference frame which can be either “I” or “P” frame. “P” frames are more efficient in terms of number of bits required compared to “I” frames, but still require more bits than “B” frames. “B” frames require the lowest number of bits compared to both “I” and “P” frames but incur computational complexity.

Thanks for your attention